Science Arts & Métiers (SAM)
is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: https://sam.ensam.eu
Handle ID: .http://hdl.handle.net/10985/9702

To cite this version :

Thomas DEVOGELE, Laurent ETIENNE, Cyril RAY - Maritime monitoring - 2013

Any correspondence concerning this service should be sent to the repository Administrator : scienceouverte@ensam.eu

PART ONE

MOBILITY APPLICATIONS

Maritime Monitoring

Thomas Devogele (Université François Rabelais de Tours, France), Laurent Etienne (Dalhousie University, Department of Industrial Engineering, NS, Canada), Cyril Ray (Naval Academy Research Institute, France)

Contents

PART ONE MOBILITY APPLICATIONS *page* 1
List of illustrations iv
List of tables v
List of contributors vi
1 Maritime Monitoring 1
1.1 Maritime Context 1
1.1.1 Maritime Traffic 1
1.1.2 Maritime Positioning Systems 3
1.2 A Monitoring System Based on Data Mining Processes 6
1.2.1 Platform, Database Model 8
1.2.2 From Raw Positions to Trajectories 10
1.2.3 Trajectory Clustering Process 13
1.2.4 Spatio-Temporal Pattern Mining 14
1.2.5 Outlier detection 18
1.3 Conclusion 20
1.4 Bibliographic Notes 21

Notes 23
References 24
Author index 27
Subject index 28

Illustrations

1.1	Ships' density map in Europe	2
1.2	Data mining and trajectories' qualification process	7
1.3	From raw data to trajectory pattern (Bay of Brest)	10
1.4	Clusters of positions and spatial pattern	16
1.5	3D spatio-temporal pattern and outlier trajectories	17
1.6	Outlier detection	19

Tables

1.1 AIS shipborne mobile equipment reporting intervals 6
1.2 Database model 9
1.3 Results for filtering process with 10 m tolerance 13

Contributors

Thomas Devogele *Université François Rabelais de Tours, Laboratoire d'informatique LI, 3 place Jean Jaurès, 41000 Blois, France, thomas.devogele@univ-tours.fr*

Laurent Etienne *Dalhousie University, Department of Industrial Engineering, 5269 Morris St., PO BOX 15000, Halifax, NS, Canada, laurent.etienne@dal.ca*

Cyril Ray *Naval Academy Research Institute, 29240 Brest Cedex 9, France, cyril.ray@ecole-navale.fr*

1
Maritime Monitoring

1.1 Maritime Context

The maritime environment still represents an unexploited potential for modelling, management and understanding of mobility data. The environment is diverse, open but partly ruled, and covers a large spectrum of ships from small sailing-boats to super tankers which generally exhibit type-related behaviours. Similarly to terrestrial or aerial domains, several real-time positioning systems, such as the *Automatic Identification System* (*AIS*), have been developed for keeping track of vessel movements. However the huge amounts of data provided by these reporting systems are rarely used for knowledge discovery. This chapter aims at discussing different aspects of maritime mobilities understanding. This chapter enables readers to, first, understand the intrinsic behaviour of maritime positioning systems and then proposes a methodology to illustrate the different steps leading to trajectory patterns for the understanding of outlier detection.

1.1.1 Maritime Traffic

The maritime environment has a huge impact on the world economy and our everyday lives. Beyond, being a space where numerous marine species live, the sea is also a place where human activities (sailing, cruising, fishing, goods transportation...) evolve and increase drastically. For example, world maritime trade of goods volume has doubled since the seventies and reached about 90% of global trade in terms of volume and 70% in terms of value. This ever increasing traffic leads to navigation difficulties and risks in coastal and crowded areas where numerous ships exhibit different movement objectives (sailing, fishing, ...) which can be conflicting. The disasters and damages caused in the event of sea collisions can pose serious threats to the environment and human lives. Such disasters and damages often lead to highly negative effects on maritime ecosystems and are threats not only for the important populations of marine protected and endangered species, but also for economic, scientific, and cultural sectors. Safety and security have therefore become a major concern, especially in Europe.

Consideration of this security issue by the International Maritime Organization (IMO) has partly evolved in the last decade from ship design, education, navigational rules (e.g. International Regulations for Preventing Collisions at Sea: $COLREGS$), to technical answers for traffic monitoring. Nowadays, ships are fitted out with almost real-time position report systems whose objective is to identify and locate vessels at distance. Figure 1.1 shows, for instance, ships' trajectories obtained through the *Automatic Identification System* (AIS) in Europe during one month.

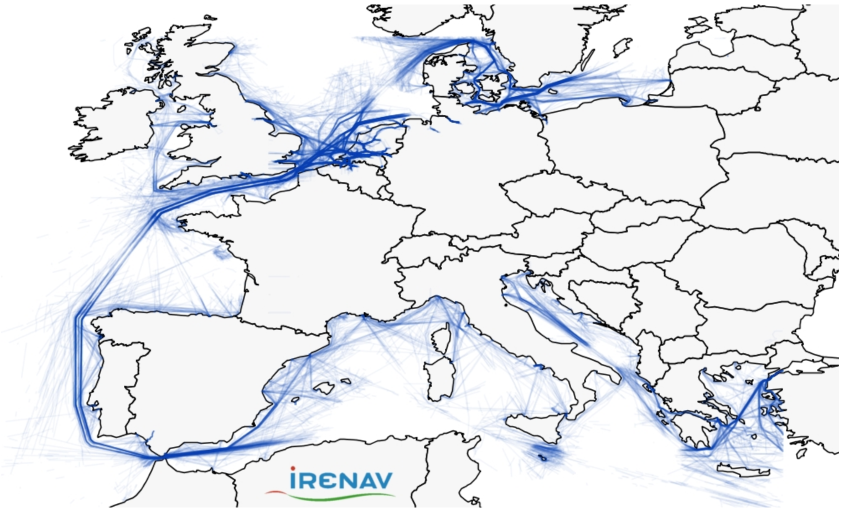

Figure 1.1 Ships' trajectories, density map in Europe during one month (AIS positions, December 2010).

The maritime environment, represented in Figure 1.1, is diverse and open, but partly ruled. Regulation is ensured by Traffic Separation Schemes (TSS) set up in order to split and regulate the traffic in crowded spaces into traffic-lanes, and by the definition of exclusion areas and Particularly Sensitive Sea Areas ($PSSA$) the ships have to avoid (e.g. biodiversity areas. Trajectories in such an open space are very typical; ships often behave similarly, in straight lines, leading to visually noticeable trends and patterns. This naturally favours the analysis of aggregated behaviours in order to detect maritime routes, dense areas, evolution of the traffic, and finally at individual levels, abnormal trajectories and locations.

1.1.2 Maritime Positioning Systems

Two of the most successful systems used in maritime navigation and positioning are the Automatic Radar Plotting Aid ($ARPA$) and the Automatic Identification System (AIS). Both are used by vessels and Vessel Traffic Services on shore (VTS) in order to facilitate navigation decisions and warn about possible collisions. Vessel traffic services also take advantage of their higher computing and networking resources to store data locally and share them at national and worldwide levels (e.g. program SafeSeaNet of the European Maritime Safety Agency).

Marine radar with automatic radar plotting aid tracks vessels using radar contacts. Radar transmitter generates very short pulses of radio waves. When the radio waves of one of these pulses encounter any obstacle, such as a ship, shore line or big sea waves, part of the radiated energy is reflected and received by the emitting radar. The reflected pulse constitutes a radio echo. The time between the pulse and the echo can be accurately measured and used to calculate the distance between the radar and the echo. The direction of the echo reflects the direction of the pulse. When a target echo appears on a radar screen, an operator plots the relative motion of the echo in order to get the target's course and speed. The maximum range of an object detected is affected by the height of the radar antenna as well as the height of the object due to the curvature of the earth. In the same way, mountainous sea lines cause blind areas, and objects behind these areas cannot be detected. Bad weather conditions can also affect significantly the effectiveness of radar tracking. Thus, any target should be acquired and confirmed in at least five of ten scans over a period of 2 minutes in order to be brought to the attention of the operator with an identifier and coordinates.

Automatic Identification System has been recently implemented and made a mandatory standard on commercial and passenger ships. This system, whose objective is to identify and locate vessels at distance, automatically broadcasts location-based information through self organised wireless communications (VHF). AIS usually integrates a transceiver system, a GPS receiver, and other navigational sensors on board, such as a gyrocompass and a rate of turn indicator. An AIS transponder runs in an autonomous and continuous mode, and regularly broadcasts a position report according to the ship's behaviour. The information is broadcast, within a range of 35 nautical miles, to surrounding ships and maritime authorities on the ground. There are two different classes of AIS that can be found on ships, search and rescue aircrafts and base stations on ground: Mandatory AIS (class A) for large vessels and low-cost AIS (class B) which has been introduced for smaller vessels. Devices from these two classes broadcast information at different time intervals (table 1.1), and at different ranges (typically 20-40 miles for class A and generally 5-10 miles for class B).

Enhanced worldwide positioning systems are emerging especially to address drawbacks of both systems which are complementary but imperfect. On one hand, $ARPA$ is useful to detect and track vessels that might not have AIS devices on-board. On the other hand, it brings limited information and cannot identify a mobile object, and its coverage include blind areas. The automatic identification system is useful to obtain more complete information, but devices are not available on all ships and data can be falsified. The most important issue that guides evolutions concerns the limited tracking range of both systems which is insufficient to follow ships engaged on international journeys. Satellite communications systems are going to be more intensively employed, in particular to enhance or replace the AIS. For instance, Long-Range Identification and Tracking ($LRIT$) reports vessel positions to their flag administration at least four times a day. Satellite-based AIS-monitoring service ($S - AIS$) uses satellite communications to broadcast AIS information. Nowadays, position reports for European coasts reach almost 1.5 million positions per day (about 72,000 ships). The ever increasing data flows provided by this evolution is going to emphasize issues on maritime data integration, fusion, filtering, processing, and analysis.

Location-based data: While radar data is limited to a tuple composed by an identifier, a position, and a related time, the automatic identification system broadcasts a wide range of richer information. Information systems on-board or in vessel traffic services generally merge AIS and radar positions into a single accurate one. When a ship is not fitted with an AIS (typically small boats), the reported information for data analysis is only limited to the aforementioned tuple. From our perspective, this does not impact the data-mining process and therefore motivate an analysis focusing on the *AIS* data more easily accessible. Transmitted *AIS* data comes from twenty-seven different messages each providing specific information either related to the behaviour of the AIS system or to ship's locations and characteristics. Positioning data defines point-based trajectories describing two-dimensional routes on the sea surface. That is, an ordered series of locations expressed in WGS84 format (latitude λ, longitude φ, time t) of a given mobile object with t indicating the timestamp of the location (λ, φ). Among all the received data, meaningful information that can be considered in a purpose of movement discovery and understanding can be classified in the three following categories:

- *Static*: MMSI number (Maritime Mobile Service Identity: a unique ID), Name, Type, International Maritime Organization Code, Call sign, Dimension.
- *Dynamic*: Position (Longitude, Latitude), Time, Speed, Heading, Course over ground (COG), Rate of turn (ROT), Navigational status.
- *Trajectory-based*: Destination, Estimated time of arrival (ETA), Draught, Dangerousness.

Quality of data is variable and depends, first, on the quality of the *AIS* device itself and the way it implements algorithms and protocols. Therefore data like coordinates and speed can be more or less accurate. Longitude and latitude are normally given in 1/10000 minute that should give 0,18m. However, considering this quality factor and intrinsic behaviour of GPS, the International Maritime Organization only considers an accuracy of 10m. The quality also depends on people on-board. Indeed, some data like MMSI, name, destination or navigational status are manually set and possibly wrong. Contextual information associated to geographic positions helps to understand ships behaviours according to space, time, destination, and ships' types although they require error-detection and filtering processes.

Space and time gaps: Time is not part of position reports as the *AIS* had been initially designed for real-time purpose only. Each received message has to be timestamped with the receiver's clock. While it communicates on a regular basis, the automatic identification system does not send these position reports continuously. Transponders broadcast data to surrounding listeners at different sampling rates according to ship's behaviours. Table 1.1 presents sampling rates for *AIS* class A. Class B devices behave on a similar way but at different sampling rates. This variation of time intervals is very specific to maritime domain and can vary from 2 seconds for a fast moving ship to several minutes when anchored.

Table 1.1 *AIS shipborne mobile equipment reporting intervals.*

Ship's dynamic conditions - AIS Class A	**Freq.**
Ship at anchor or moored and not moving faster than 3 knots	3 m
Ship at anchor or moored and moving faster than 3 knots	10 s
Speed between 0 and 14 knots	10 s
Speed between 0 and 14 knots and changing course	$3\ \frac{1}{3}$ s
Speed between 14 and 23 knots	6 s
Speed between 14 and 23 knots and changing course	2 s
Speed over 23 knots	2 s
Speed over 23 knots and changing course	2 s

The range covered by all VTS on shore is limited and coverage areas might not overlap everywhere. In such a context, the observation of the maritime traffic at a given time lead to a partial view due to space and time gaps. These received positions will mostly not correspond to the selected times for snapshots analysis (*e.g.* a ship communicated its position 10 seconds before the analysis time). This implies to consider time intervals and the definition of trajectories for a successful analysis and understanding of the ships' behaviours. Let us note that these large and variable gaps between two position reports will affect significantly the way trajectories can be computed.

1.2 A Monitoring System Based on Data Mining Processes

The increase of maritime location-based information brings opportunities for knowledge discovery on movement behaviours at sea over a long period of time. This section shows how maritime data can be processed and analysed in order to qualify a given position or trajectory with computed patterns. This process allows for instance to detect outliers including real-time traffic monitoring. It is based on data-mining principles presented in other chapters, especially Chapter 6. The methodology postulates that *normal* moving objects following a same itinerary at sea behaves in a similar optimised way. Such a behaviour illustrated in figure 1.1 helps to compute accurate trajectory patterns.

Figure 1.2 presents the functional process used to extract spatio-temporal patterns from spatio-temporal databases and qualify ship positions and trajectories. First, an acquisition step (Step 1 in figure 1.2) integrates AIS raw data from several monitoring systems into a structured spatio-temporal database ($STDB$). In this database, zones of interest (ZOI) define either an origin or a destination of a trip. Each identified ZOI is associated to its surface and linked to its neighbours (and stored in the spatio-temporal database). Then trajectories are clustered (Step 3) according to their itineraries in order to obtain Homogeneous Groups of Trajectories (HGT). A statistical analysis of these clusters gives the median trajectory of each cluster and spatio-temporal intervals around them (Step 4). Median trajectories and intervals are combined together to define the spatio-temporal pattern of $HGTs$. These patterns are stored in a knowledge database (Step 5). They can be used either for geovisual analyses or to qualify in real-time ship positions and trajectories (Step 6).

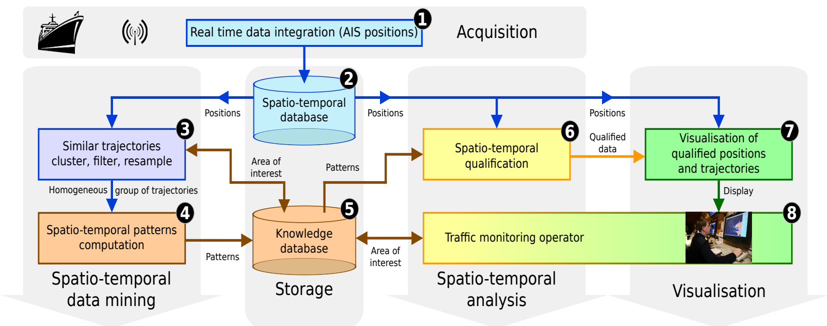

Figure 1.2 Data mining and trajectories' qualification process.

This functional process has been experimented and used in different contexts: real-time tracking of sailing races and maritime navigation in the coastal area of Brest: Processing and analysis of AIS raw data from Aegean, North and East China seas, and from aggregated real-time data flows from $NATO$ countries. A maritime case study based on *passenger* *ships* in bay of Brest, France illustrates, throughout this chapter, this qualification process for safety purpose (a sample dataset is available at ChoroChronos repository, http://www.chorochronos.org).

1.2.1 Platform, Database Model

This functional process (figure 1.2) relies on a generic and scalable information system that has been designed for real-time monitoring and spatio-temporal analysis of different types of moving objects at sea. So far, the underlying platform developed is a Java-based computing system based on a *PostgreSQL/PostGIS* spatial database for data manipulation and storage. It has been designed with four tiers client-server architecture, and organised through a distributed data and processing model. The information system is based on different functions:

- real-time integration of positioning information (figure 1.2, Step 1),
- spatio-temporal data mining (figure 1.2, Steps 3-5),
- spatio-temporal analysis (figure 1.2, Step 6),
- web-based visualisation (figure 1.2, Step 7).

The data model set up in the *PostGIS* database relies on the aforementioned classification of *AIS* messages: static, dynamic and trajectory-based (table 1.2). Table labelled *AISPositions* stores all the dynamic position reports of ships. Table *AISShips* contains the static information, especially ship's type which can be used later to cluster trajectories of similar ships (cargo, passenger ships, sailing ships...). Table *AISTrips* is used to store ships' trip based on information such as its destination and the type of goods it is carrying. In addition to these tables that are containing raw information, some derived data can be added to the database. Table *Trajectories* is obtained from positions of the table *AISPositions* and from *AISTrips* in order to link position reports of a same ship together and to reconstruct its path (table 1.2, field *trajectories.shape*). As table *AISTrips* gives information about ships' destinations, these destinations can be extracted as zones of interest (*ZOI*) and stored in a new table *Zones*. The zones of interest can also be manually defined by an operator according to various criteria such as regulations (waiting areas, traffic channels, restricted areas), geography (obstacles, isthmuses, straits, inlets), and economy (shops, loading sites, ports, fishing areas). These zones of interest, represented as spatial zones, can be connected together to define a *zone graph* in order to analyse ships' mobility and describe their itinerary (Table *Itineraries*).

Table 1.2 *Database model.*

Table	Description
	Data provided by AIS
AISPositions	Position reports of each ship with additional dynamic information MMSI (numeric), Time (timestamp), Heading (numeric), Speed (numeric), COG (numeric), ROT (numeric), Coordinates (geometry), Status (text)
AISShips	Static information about ships MMSI, OMI Number (numeric), Name (text), Callsign (text), Type (text), Length (numeric), Width (numeric)
AISTrips	Trajectory-based information MMSI, Draught (numeric), Danger (boolean), Destination (text), ETA (timestamp), Reported Time (timestamp)
	Derived data added to the model
Trajectories	Trajectories extracted from raw data MMSI, Beginning Time (timestamp), End Time (timestamp), Shape (geometry)
Zones	Zones of interest (ZOI) ZID (numeric), Name (text), Shape (geometry)
Itineraries	Itineraries between ZOI IID (numeric), Start Zone ID (numeric), End Zone ID (numeric)

For richer analysis, taking geographic information into account might also be of interest. The database could therefore include a large set of tables obtained from official $S-57$ vector charts that contain different kind of objects useful for spatial analysis:

- Points of interest: buoys, shipwrecks, containers at sea, ...
- Lines of interest: coastlines, path, channels, crossing lines, ...
- Zones of interest: oil spill, ports, restricted areas, $PSSA$, ...

The zone graph of the bay of Brest is illustrated on figure 1.3.b. The numerous dots shown in figure 1.3.a represent positions of ships. An itinerary (I) is an arc between two zones of the graph. Parts c and d of figure 1.3 illustrating trajectory patterns will be presented later in sections 1.2.3 and 1.2.4.

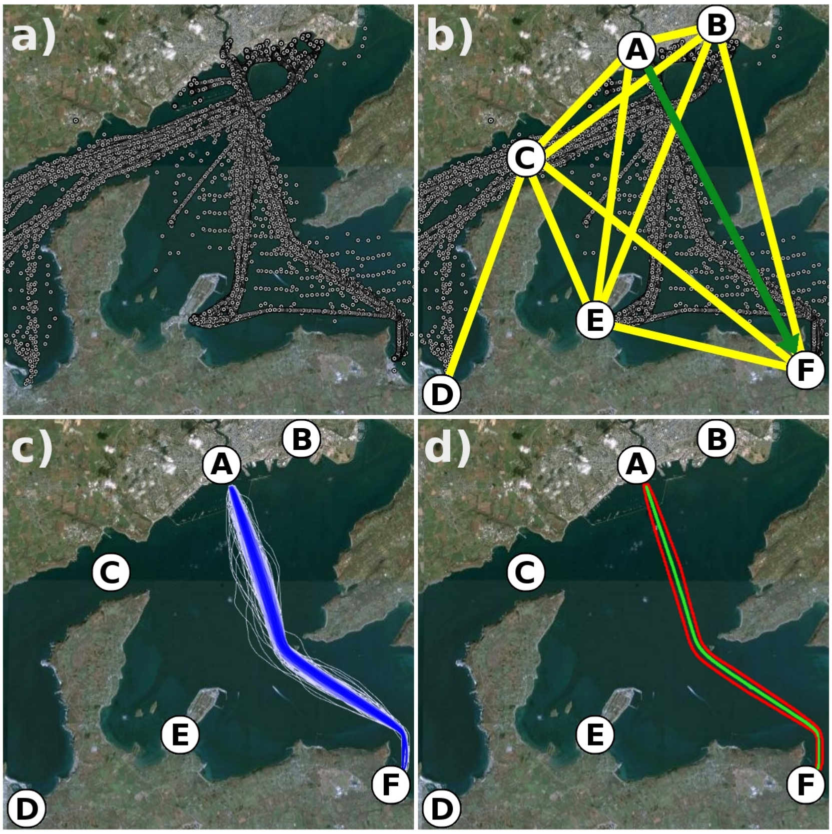

Figure 1.3 From raw data to trajectory pattern (Bay of Brest).

1.2.2 From Raw Positions to Trajectories

As shown in figure 1.3.a, the numerous position reports of ships can be put together in order to build a trajectory and address point-based query limits. Point-based queries (strictly based on raw positions) exhibit two limits. First, a computing limit as point-based spatial queries are very expensive in terms of computing cost. Second, it reaches a spatial limit as queries are applied on reported locations provided by the *AIS* (a ship passing through a narrow restricted area can report positions on both sides, due to *AIS* behaviour and sampling frequency, even if the trajectory of the ship crosses the zone). Therefore it is difficult to identify whether a ship went through a narrow passage, entered a restricted area or computed exact minimal distances to the coast (this requires interpolation and additional computing costs).

Trajectory features are required to query more correctly and efficiently the AIS database. Further, it allows for distance computation based on polylines instead of raw positions, route definitions, trajectory comparisons and clear identification of passage through an area or a line. Due to the computing limit, number of positions for each trajectory must be reduced using a filtering algorithm in order to apply spatial operators and functions to efficiently answer end-users questions. This trajectories production stage is located between Step 2 and 3 of the data mining and qualification process (figure 1.2).

Many approaches can be considered to define what is a maritime trajectory and build such trajectories from a sequence of AIS positions. Let's consider the time ordered sequence of all AIS positions of a given ship defined by $S = \{p_0, ..., p_n\}$. A trajectory T of this ship can be defined as a sub-sequence of S so that $T \subset S \wedge T = (p_b, ..., p_j, ..., p_e)$ where p_b stands for the beginning position of the trajectory and p_e for the ending one.

The main matter consists in selecting the beginning and ending positions from S in order to create a set of trajectories. These particular positions (considered as stops) can be identified by the mobile object cinematic (e.g. null speed), its spatial position (inside a zone of interest) or the position report sampling rate (transmission gaps). As the position reports from the AIS itself are not regular and depend on the ship's behaviour (table 1.1), a simple time and spatial threshold might not be sufficient to properly detect gaps defining the beginning and ending positions and split sequences of raw positions into trajectories. So, a dynamic spatial (δs) and temporal (δt) thresholds should be derived from the enriched information provided by the AIS which contains heading H_p, speed S_p, acceleration A_p, and rate of turn R_p indications. Such an approach can rely on the number of missed frame(s) allowed (n_{mf}) and the reporting intervals expected by the AIS device on-board (table 1.1) to define the time (δt) and spatial (δs) thresholds. The next position of a trajectory should be transmitted within δt and should be located within a maximum δs distance. Otherwise, the last position is considered as a stop and future positions of the sequence S will be associated to a new trajectory.

Another way to define these stops within a sequence of positions is to rely on zones of interest which can be identified in cartographic information or manually defined by an expert (cf. section 1.2.1). This inevitably changes the semantic of the trajectory with respect to the previous method. However, such an approach suits better to the analysis of maritime mobilities as ships always have a small number of well-defined origins and destinations (harbour, mooring or waiting area). For a more automatic process, such areas can be also created automatically using a density analysis. In this context, a beginning position of a trajectory is a position which is inside a zone Z and whose next position of the trajectory is outside this zone. A ending position of a trajectory is a position which is inside a zone Z and whose previous position of the trajectory is outside this zone. Therefore, the sequence S of position of a given ship can then be split into a subset of trajectories $\Gamma = \{T_0, ..., T_N\}$ such as $\Gamma \subseteq S$.

Once the positions are assigned to trajectories, a filtering process selects the key positions of a given trajectory. A position is considered as a key position when either the speed or the direction changes significantly. The other positions can be removed.

The algorithm initially introduced by Douglas and Peuker in 1973 is relevant as it performs well on typical straight trajectories of vessels. The principles of the original algorithm are as follows. The start and end points of a given polyline are connected by a straight line segment. Perpendicular offsets for all intervening end points of segments are then calculated from this segment, and the point with the highest offset is identified. If the offset of this point is less than the tolerance distance, then the straight line segment is considered adequate for representing the line in a simplified form. Otherwise, this point is selected, and the line is subdivided at this point of maximum offset. The selection procedure is then recursively applied to the two parts of the polyline until the tolerance criteria is satisfied. Selected points are finally chained to produce a simplified line.

This simplification algorithm for trajectory filtering could be adapted in order to be more efficient. Conversely to Meratnia and By in 2004 who used Euclidean Distance between points at a same time, the Haversine distance can be used. This distance is the shortest distance (d_s) between two points measured along a path on the surface of a sphere. The perpendicular distance is therefore derived as a spatio-temporal distance d_{ST} and is as follows:

$d_{ST}(T_i, T_j, t) = d_s(p_i(t) - p_j(t))$

The spatio-temporal distances between position p_i of the trajectory T_j, and position p'_i of the interpolated trajectory T'_j taken at a same time (relative time from the departure) are computed. Let us note that these

Table 1.3 *Results for filtering process with 10 m tolerance.*

Vessel	Trajectory duration	% of position kept	% of length kept (km)
Bindy	28m 01s	14.0% (32/229)	99.91% (11.284/11.294)
Port pilot boat	1h 07m 36s	21.7% (122/562)	99.82% (24.846/24.892)
AB Valencia	7h 04m 20s	12.0% (279/2316)	99.98% (175.07/175.109)

spatio-temporal distances are influenced by the speed and the direction of the mobile object. A tolerance distance should be defined appropriately. According to the GPS position accuracy, a tolerance of 10 meters is acceptable.

In order to exemplify this filtering process, three vessel trajectories have been selected for illustration purpose. The first trajectory concerns a passenger boat, called *Bindy* whose trajectory is smooth and speed is regular. The second trajectory is the one of a port pilot ship in the harbour of La Rochelle. This trajectory is very sinuous, and several loops appeared. The third trajectory is composed of long straight polylines made by the cargo ship *AB Valencia*.

Table 1.3 summarises the filtering result. One can note that their lengths are very close. This leads to a filtering process where more than 80% of the received positions can be filtered. The performance of the filtering process is likely to increase for large ships, and decrease for small ships due to the intrinsic characteristics of their navigation.

1.2.3 Trajectory Clustering Process

Once the trajectory concept is defined, different trajectory clustering techniques can be used to get homogeneous groups of trajectories. Some of them are presented in Chapter 6. Another technique based on the zone graph and itineraries can be used to extract clusters from trajectories following the same itinerary I. This set is called a homogeneous group of trajectories (HGT).

The first selection criterion of this approach is based on static information such as the type of mobile objects; This information is provided by AIS messages (table 1.2). The second selection criterion is a geographical one. The first position of the trajectory (p_b) must be the only one within the departure zone (Z_D) of the itinerary, and the last position of the trajectory (p_e) must be the only one within the arrival zone (Z_A) of the itinerary. Taking into account the frequency of trajectory samples and the speed of the mobile object, trajectories that cross a zone of the graph should have at least one position within this zone. The last selection criterion used is time. Some moving objects can follow this itinerary periodically. These different trajectories can be distinguished using a time interval. Finally, the trajectory should not intersect any other zone of the graph G_Z that does not belong to the itinerary I. All valid trajectories previously extracted from the $STDB$ compose the HGT to be analysed.

Figure 1.3.c illustrates the extraction of the HGT of 500 passenger ships' trajectories following the itinerary between Brest and Naval Academy (arc A-F of G_Z). Some density differences can be noticed on this HGT. This HGT highlights the outlier trajectories represented in light blue (outside the deep blue dense area).

1.2.4 Spatio-Temporal Pattern Mining

Once the HGT clusters have been extracted and filtered, the next step aims at defining the pattern followed by most trajectories of each HGT. The main matter of this mining task is to deduce the median trajectory followed by the HGT and the spatial and temporal density distribution. Studies on several trajectory clusters showed that this data does not belong to any particular statistical distribution. Gaps between mean and median values are important. Density around these values change frequently. For example, for time dimension, it's easier for mobile objects to arrive late than early. For this kind of ordered set of data in descriptive statics, box plot series are very useful to describe the evolution of data according times. Box Plot, proposed by John Tukey in 1977, graphically describes groups of numerical data through five important sample percentiles :

- the sample minimum (smallest observation),
- the lower quartile or the 1st decile,
- the median,
- the upper quartile or the 9th decile,

- the sample maximum (largest observation).

In our maritime context, data lower than the first decile or upper than the 9th decile are considered as outliers. The idea is to enhance box plot series to produce 2D plus time patterns. Each pattern summarises a cluster of trajectories (HGT) thanks to the median value, and the symmetry and dispersion of the data set.

First of all, a synthetic median trajectory (T_m) can be computed using an iterative refinement technique similar to the k-means algorithm. A trajectory from the HGT is chosen as initial T_m. T_m is an ordered set of positions: P_{m_i}. To optimize this algorithm, a trajectory with length and time duration close to median length and median time duration have to be chosen as initial T_m. Then, all positions of each trajectory of this HGT are assigned to one position of T_m using a matching process. Amongst existing algorithms, Dynamic Time Warping (DTW) or Fréchet matching can be employed. They can align trajectories' positions in order to minimise the sum of the spatial distances between matched positions of two trajectories (DTW), or minimise the maximum distance between matched positions (Fréchet). They also take into account the temporal order of the positions of trajectories. Figure 1.4 illustrates the clusters of matched positions ($\mathcal{C}m_{p_i}$) between positions of trajectories of the HGT and the P_{m_i} in yellow. Coloured lines show links between matched positions.

Once every position is matched, the coordinate and the timestamp of P_{m_i} are updated, by computation of median X ($\tilde{X}$), median Y ($\tilde{Y}$), and median timestamp ($\tilde{t}$). A medoid approach is also possible but requires more time for similar results. Assignment and update steps are repeated until the distance (Fréchet distance or average distance) between two consecutive points reaches a minimal threshold value.

As the studied mobile objects move in an open area, some of them can move away from the main trajectory. *Normal* slight temporal or spatial deviations must be distinguished from outliers. Two channels are computed to distinguish the spatio-temporal outliers. First, the spatial channel is defined. Once the median trajectory is computed, a statistical density analysis can be performed on every cluster of matched positions ($\mathcal{C}m_{p_i}$). These clusters are split into two subsets of positions $\mathcal{L}_{p_i}$ (left sided) and $\mathcal{R}_{p_i}$ (right sided) according to their side to the median position P_{m_i} using the P_{m_i} heading. Then, spatial distances between positions from $\mathcal{L}_{p_i}$ and the P_{m_i} are computed. After a statistical analysis, the ninth decile is chosen as left limit of the channel for this $\mathcal{C}m_{p_i}$. The same

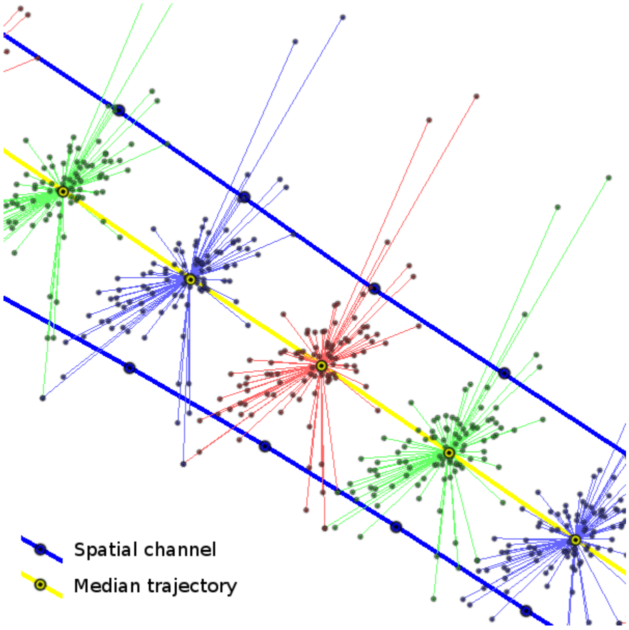

Figure 1.4 Clusters of positions and spatial pattern.

process is computed to define the right limit of $\mathcal{C}m_{p_i}$. The left (right) limits are linked according to the time to define the left (right) limit of the spatial channel. Figure 1.4 presents the limits of the spatial channel in blue. Some positions are visually outside this channel and can be defined as outliers. In the same way, the temporal channel is defined. Positions of $\mathcal{C}m_{p_i}$ inside the spatial channel are split into two subsets, late sided and early sided, according to the difference between relative timestamps of positions and on the median matched position. The early and the late limits are computed to define the temporal channel of each $\mathcal{C}m_{p_i}$. Positions outside the spatial channel are not taken into account because these parts of trajectories including these positions could be shortcuts or detours. Spatial and temporal channels at each relative time can be combined to create the spatio-temporal channel which is then stored in the knowledge database. Figure 1.3.d illustrate the spatio-temporal channel of the HGT (figure 1.3.c) extracted from zone A to F of the zone graph (figure 1.3.b). The spatial and temporal widths change. For example, for the straight part of the pattern, the spatial width is bigger than the curved part's one.

The spatio-temporal pattern defines five different zones (usual position zone, right outlier zone, left outlier zone, late outlier zone, and early outlier zone) for each relative time. This spatio-temporal pattern (median trajectory plus spatio-temporal channel) is a 2D+t enhancement of the box plot concept. It can be illustrated in 3D using the Z axis to represent the relative time as shown in Figure 1.5. The median trajectory is plotted in yellow, the usual 3D zones are the green boxes defined for some key positions of median trajectory linked together. The early limits of the spatio-temporal channel are outlined in green and the late limits in red. Two examples of outliers trajectories (cyan and purple lines) getting out of the spatio-temporal channel are presented in this figure. The purple one includes late outlier position at the end of the trajectory. The cyan one includes right outlier at the beginning of the trajectory. This spatio-temporal pattern must be computed for each HGT. As new positions are frequently acquired by the system, this spatio-temporal channel could be improved by updating it periodically.

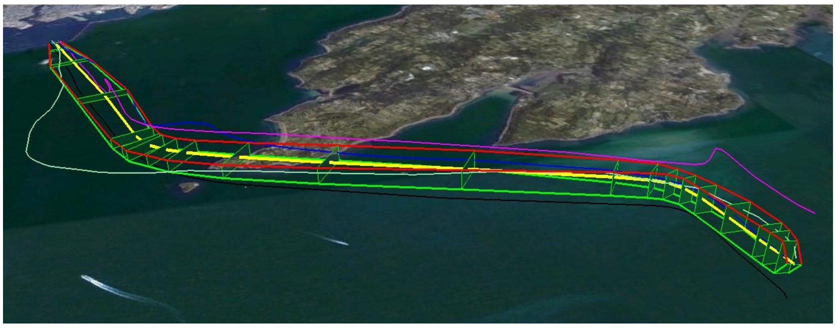

Figure 1.5 3D spatio-temporal pattern of an itinerary and outlier trajectories.

Quality of the set of patterns depends on the precision of the ZOI graph and the set of mobile object types. This quality could be verified if the spatial and temporal distributions of positions of each $\mathcal{C}m_{p_i}$ is unimodal. If several modes appear, a new analysis can be carried out to split the set of mobile objects according to types or to add new ZOI in the graph.

1.2.5 Outlier detection

For each cluster, the associated spatio-temporal pattern splits the set of trajectory positions in the outlier position group and the usual position group. For a new vessel position, this knowledge data could be useful to detect and to qualify this position. Therefore, this section suggests to combine the knowledge database and the production database to obtain an inductive database and to detect the outlier positions in real-time. Let's consider a new position p received. The position qualification process is decomposed into three steps (illustrated in Figure 1.6):

- trajectory extraction from the last ZOI encountered by the mobile object to p,
- matching process between this trajectory and the median trajectories of a pattern,
- spatio-temporal comparison between p and selected pattern.

In the first step, database is queried to select the start position from the trajectory. This position is the last one of the mobile object inside the surface of one ZOI. Positions between p and this departure position are timestamp-ordered to define a trajectory path. This last one does not link two ZOIs consequently, it is called a partial trajectory (T_p). In Figure 1.6, the last ZOI is A and the start position is (b). The partial trajectory is the blue polyline. The second step must match T_p with part of a median trajectory. This matching can be done according to:

- the type of the moving object,
- the geometry of the partial trajectory,
- the set of median trajectories from the departure ZOI,
- information about the course of the moving object to destination.

Unfortunately, information about the destination is often false or unknown, so only the type of vessel and geometries properties can be used. In order to match two linear geometries, the Fréchet discrete distance is selected as DTW does not allow partial matching processes. Fréchet distance gives the maximal distance between two lines. The Fréchet discrete distance applied to two discrete trajectories (ordered set of points) can match trajectories together preserving order of theirs points. Alt and Godau (1995) demonstrate the advantage of this measure. Devogele in 2002 proposes to enhance this distance in order to compute the distance between a line and an homologous part of another line. This partial discrete Fréchet distance ($dPdF$) is very useful to match a trajectory

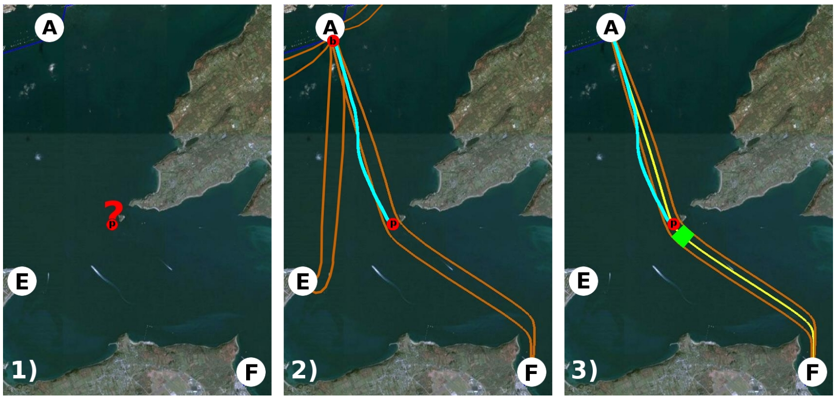

Figure 1.6 Outlier detection.

where only the departure is known. Thanks to this $dPdF$, the distance between T_p and median trajectories from the same departure ZOI can be computed. Only the spatio-temporal patterns for the same type of this object are taken into account. T_p can be partially matched with one median trajectory ($\tilde{T}$) where $dPdF(T_p;\tilde{T})$ is lower than the $dPdF$ with other median trajectories plus a threshold. $dPdF(T_p;\tilde{T})$ must also be less than a maximal value. In the example, the distance between T_p and two median trajectories (from ZOI A to E and from A to F) are computed. The second distance is the lowest, so T_p is matched with median trajectories from A to F.

Finally, the position p could be qualified according to the selected pattern. The relative time of p from departure ZOI is employed to infer the spatio-temporal channel from the knowledge database. The 3D channel is cut at this timestamp and the space is split into five areas (right, left, usual, late and early). Qualification of p is given by the area that contains p. For example, the spatial channel of the matched pattern is limited with brown lines and the usual area at the relative time of p is the green area. Position p is an outlier and is located in the late area, so this object can be spatially qualified as "inside the channel" but temporally as "running behind schedule". Such real-time analysis methods can be used to predict the destination and time of arrival of the ship once an itinerary has been matched, and if the position is normal. The destination prediction can be higher than 90%. In the same way, confidence interval of time of arrival could be the width of temporal channel at the arrival.

1.3 Conclusion

Maritime environment represents an increasing potential in terms of modelling, management and understanding of mobility data. The environment is typical and recently several real-time positioning systems, such as the *Automatic Identification System* (*AIS*), have been developed for keeping track of vessel movements. This chapter outlines different aspects of maritime mobilities understanding through pattern discovery and analysis of ships' trajectories. Underlying issues concern in particular trajectories modelling problems, trajectory querying and simplification, similarity functions, classification and clustering algorithms, and knowledge discovery (trends, unusual behaviours, and event detection).

Assuming that moving objects at sea that are following the same itinerary behave in a similar way (considered as the normality), this chapter illustrates the different steps leading to outliers's detection. The suggested methodology considers several steps. First, the data flow provided by the automatic identification systems is managed in structured spatio-temporal databases. Then, data mining processes are used to extract trajectories (vessels of the same type) and spatio-temporal patterns between two zones of interests (an origin, a destination). Each pattern includes a median trajectory and a spatio-temporal channel that describes the dispersion of the set of trajectories. Such trajectory patterns are meaningful to understand maritime traffic and detect outlier positions in real-time. Indeed, each new position (partial trajectory) can be spatially and temporally qualified according to spatial and temporal criteria. For end-users monitoring maritime traffic, such real-time qualification of positions and trajectories tied with triggers automatically executed when a new outlier is detected, and adapted geovisualisation process are essential for safety purposes.

While complete, the suggested methodology still leaves aside several additional challenges. First, cartographic information and environmental data such as currents, tides, and winds that affect ships' movements could be taken into account for further improvements. Many other algorithmic approaches for trajectory representation and reconstruction can be considered for other knowledge discovery objectives. Interactive and adaptive geovisualisation is also of interest. Another challenge concerns new itineraries. Many factors can influence ship's behaviour leading to the apparition of new itineraries. The proposed approach handles such regular trajectories as outliers. An adaptive process should be therefore considered in order to detect a new pattern and possibility remove an outdated one. Finally, the approach described could be applied or extended to other kind of moving objects evolving in open spaces especially those having 3-dimensional trajectories (e.g. underwater vehicles or planes that behave quite similarly to ships).

1.4 Bibliographic Notes

Several maritime projects worked to enhance the tracking and monitoring of vessels. This is portrayed for example in $MarNIS$ MarNIS (2009). These monitoring systems use $ARPA$ and AIS sensors as input. Bole et al. (2012) describes $ARPA$ system in detail. In a similar way, the Association of Marine Aids to Navigation and Lighthouse Authorities describes the AIS in IALA (2004). These new tracking and monitoring systems are parts of e-Navigation defined by the International Maritime Organization in IMO (2008). e-Navigation relies on Electronic Navigation Chart (ENC) defined by the International Hydrographic Organization in IHO (2000).

If reader needs additional information about some special technical points of this chapter, several articles can be read. For the filtering part Meratnia and de By (2004) serves as the base for the filtering process presented in this chapter. For the similarity measure between trajectories, Fréchet Distance has been selected. Alt and Godau (1995) explains why this measure is better for this kind of data. Devogele (2002) describes the algorithm for discrete partial Fréchet distance. Matching process based on Dynamic Time Warping is also possible Sakoe and Chiba (1978). Results are very similar but this later process can align only whole trajectories. Some details about our architecture are introduced in Bertrand et al. (2007). Finally, Etienne et al. (2012) details the clustering process and the spatio-temporal pattern based on Box plot. This representation is defined in Tukey (1977).

Notes

References

Alt, H., and Godau, M. 1995. Computing the Fréchet Distance Between two Polygonal Curves. *International Journal of Computational Geometry and Applications*, **5**(1), 75–91.

Bertrand, F., Bouju, A., Claramunt, C., T., Devogele, and Ray, C. 2007. Web Architecture for Monitoring and Visualizing Mobile Objects in Maritime Contexts. Pages 94–105 of: *Proceedings of the 7th International Symposium on Web and Wireless Geographical Information Systems (W2GIS), LNCS 4857*. Springer-Verlag.

Bole, A., Dineley, W., and Wall, A. 2012. *Radar And ARPA Manual: Radar And Target Tracking For Professional Mariners, Yachtsmen And Users Of Marine Radar, Third Edition*. Elsevier.

Devogele, T. 2002. A New Merging Process for Data Integration Based on the Discrete Fréchet Distance. Pages 167–181 of: *Proceedings of the 10th International Symposium on Spatial Data Handling (SDH)*. Springer-Verlag.

Etienne, L., Devogele, T., and Bouju, A. 2012. *Spatio-Temporal Trajectory Analysis of Mobile Objects Following the Same Itinerary*. In Advances in Geo-Spatial Information Science, Shi, Goodchild, Lees, Leung (Eds.), ISPRS Orange Book series, CRC Press. Chap. Modeling Space and Time, pages 47–58.

IALA, International Association Of Marine Aids To Navigation & Lighthouse Authorities. 2004. *The Automatic Identification System (AIS), Volume 1, Part I, Operational Issues, Edition 1.3*. IALA Guideline No. 1028.

IHO, International Hydrographic Organization. 2000. *Transfer standard for digital Hydrographic data, Edition 3.1*. Special Publication No. 57.

IMO, International Maritime Organization. 2008. *Development of an E-Navigation Strategy*. Reports of Sub-Committee on Safety of Navigation, 54th session.

MarNIS, Maritime Navigation Information Services. 2009. *Final Report*. MarNIS/D-MT-15/Final Report/DVS/05062009/version 2.0.

Meratnia, N., and de By, R.A. 2004. Spatio-temporal compression techniques for moving point objects. Pages 765–782 of: *Proceedings of the 9th international conference on extending database technology, LNCS 2992*. Springer-Verlag.

Sakoe, H., and Chiba, S. 1978. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, **26**(1), 43–49.

Tukey, J.W. 1977. *Exploratory Data Analysis*. Addison-Wesley.

Author index

Ray, C., 1
Devogele, T., 1
Etienne, , 1

Subject index

Maritime positioning, 5
  Automatic Identification System, 5
  Radar, 5

Maritime trajectories, 12

Outlier detection, 20
  Fréchet distance, 20
  Trajectory matching process, 20

Spatio-temporal pattern, 16
  Median Trajectory, 17
  Spatio-temporal Box Plot, 17